

**MORRISON FOERSTER**

2100 L STREET, NW  
SUITE 900  
WASHINGTON  
DC 20037

TELEPHONE: 202.887.1500

FACSIMILE: 202.887.0763

WWW.MOFO.COM

MORRISON & FOERSTER LLP

AUSTIN, BERLIN, BOSTON, BRUSSELS,  
DENVER, HONG KONG, LONDON, LOS  
ANGELES, MIAMI, NEW YORK, PALO  
ALTO, SAN DIEGO, SAN FRANCISCO,  
SHANGHAI, SINGAPORE, TOKYO,  
WASHINGTON, D.C.

Writer's Direct Contact

+1 (202) 887-6945

CMHomer@mofo.com

November 13, 2024

**Via ECF**

Hon. Ona T. Wang  
United States District Court  
Southern District of New York  
500 Pearl Street  
New York, NY 10007-1312

Re: *The New York Times Co. v. Microsoft Corporation, et al.*, No. 23-cv-11195-SHS  
*Daily News LP, et al. v. Microsoft Corporation, et al.*, No. 24-cv-3285-SHS  
*Center for Investigative Reporting, Inc. v. OpenAI, Inc., et al.*, No. 24-cv-4827-SHS

Dear Judge Wang:

OpenAI submits this status report regarding custodial statistics in response to the Court's order (Dkt. 304) after the October 30, 2024 status conference regarding Plaintiffs' motion to compel additional custodians (Dkt. 204). OpenAI has already agreed to search and produce documents from 26 custodians; Plaintiffs seek 8 more. The 8 disputed custodians are: Ilya Sutskever; James Betker, William Fedus, Daniel Kokotajlo, Brad Lightcap, Ryan Lowe, William Saunders, and Wojciech Zaremba. Collectively, these 8 additional custodians have over 170 gigabytes of data comprising nearly 2 million documents. The below chart reflects the custodial data and documents segregated for each of the 8 disputed custodians.

| Custodian         | Size (GB) | Files     |
|-------------------|-----------|-----------|
| Sutskever, Ilya   | 49.35     | 209,899   |
| Betker, James     | 4.58      | 35,090    |
| Fedus, William    | 30.40     | 1,093,452 |
| Kokotajlo, Daniel | 1.37      | 17,028    |
| Lightcap, Brad    | 49.23     | 293,125   |
| Lowe, Ryan        | 3.80      | 56,357    |
| Saunders, William | 3.17      | 76,336    |
| Zaremba, Wojciech | 28.18     | 141,748   |
| <i>Total</i>      | 170.08    | 1,923,025 |

**MORRISON FÖERSTER**

Hon. Ona T. Wang  
November 13, 2024  
Page 2

The parties have also been negotiating custodial search terms. At present, Plaintiffs’ proposed terms, across the *existing* custodians, would require OpenAI to review over 630,000 documents – exceeding the Court’s target.<sup>1</sup> See 10/30/2024 Hearing Tr. at 74:14-75:9 (COURT: “I’m going to impose right now a ceiling of 500,000 hits.”). Adding unique hits for the 8 disputed custodians would inflate that number further—bringing the total number of documents to review to approximately 750,000. These competing hit reports are summarized below.

| Custodians  | OpenAI Proposed Terms |                        | Plaintiffs’ Proposed Terms |                        |
|---|-----------------------|------------------------|----------------------------|------------------------|
|   | Size (GB)             | Files (incl. families) | Size (GB)                  | Files (incl. families) |
| 26 existing custodians - total hits               | 243.62                | 169,725                | 593.95                     | 634,112                |
| 8 disputed custodians - total hits                | 238.38                | 121,082                | 556.41                     | 352,989                |
| 8 disputed custodians - de-duplicated hits        | 114.56                | 54,014                 | 175.93                     | 122,926                |
| 8 disputed custodians - duplication %             |                       | 55.39%                 |                            | 65.18%                 |
| <b>34 custodians - total hits (de-duplicated)</b> | 358.2                 | 223,739                | 769.9                      | 757,038                |

Included in the above chart is OpenAI’s estimate of the duplication rate between the 8 disputed custodians and the existing 26 custodians. For OpenAI’s proposed terms, the duplication rate is 55%. For Plaintiffs’ proposed terms, the rate is even higher: 65%. Under Plaintiffs’ proposed search terms, the 8 additional custodians would add over 122,000 hits, requiring OpenAI to review more than 750,000 documents after de-duplication, or 50% more than the Court’s target. Under OpenAI’s proposed search terms, the additional 8 custodians would add over 54,000 hits, and require OpenAI to review over 223,000 documents. As additional context, for OpenAI’s review to date, only around 25% of documents hitting on OpenAI’s proposed search terms have been responsive.

---

<sup>1</sup> This letter references Plaintiffs’ October 8, 2024 search term proposal. Plaintiffs sent OpenAI a new proposal after business hours on November 13, 2024. OpenAI was not able to generate a search term hit report for those terms before filing this letter.

## **HORRISON FOERSTER**

Hon. Ona T. Wang  
November 13, 2024  
Page 3

Pursuant to the Court's instructions regarding prioritizing discovery at the October 30 hearing, and taking into account the substantial volume of hits, high custodial duplication rates, and low document responsiveness rates to date, OpenAI will continue to meet and confer with Plaintiffs in an attempt to reach agreement regarding both the 8 disputed custodians and a final set of targeted search terms.

Respectfully submitted,



Carolyn M. Homer